

Neural network integration during the perception of in-group and out-group members

Greven, Inez M; Ramsey, Richard

Neuropsychologia

DOI:

[10.1016/j.neuropsychologia.2017.09.036](https://doi.org/10.1016/j.neuropsychologia.2017.09.036)

Published: 01/11/2017

Peer reviewed version

[Cyswllt i'r cyhoeddiad / Link to publication](https://doi.org/10.1016/j.neuropsychologia.2017.09.036)

Dyfyniad o'r fersiwn a gyhoeddwyd / Citation for published version (APA):

Greven, I. M., & Ramsey, R. (2017). Neural network integration during the perception of in-group and out-group members. *Neuropsychologia*, 106(November), 225-235.
<https://doi.org/10.1016/j.neuropsychologia.2017.09.036>

Hawliau Cyffredinol / General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Journal: Neuropsychologia

Title:

Neural network integration during the perception of in-group and out-group members

Running title:

Group bias modulation of person perception

Inez M. Greven & Richard Ramsey

Wales Institute for Cognitive Neuroscience, School of Psychology, Bangor University,
Bangor, Gwynedd, Wales, LL57 2AS, United Kingdom

Corresponding author: r.ramsey@bangor.ac.uk, School of Psychology, Bangor University,
Bangor, Gwynedd, Wales, LL57 2AS, United Kingdom

Abstract

Group biases guide social interactions by promoting in-group favouritism, but the neural mechanisms underpinning group biases remain unclear. While neuroscience research has shown that distributed brain circuits are associated with seeing in-group and out-group members as “us” and “them”, it is less clear how these networks exchange signals. This fMRI study uses functional connectivity analyses to investigate the contribution of functional integration to group bias modulation of person perception. Participants were assigned to an arbitrary group and during scanning they observed bodies of in-group or out-group members that cued the recall of positive or negative social knowledge. The results showed that functional coupling between perceptual and cognitive neural networks is tuned to particular combinations of group membership and social knowledge valence. Specifically, coupling between body perception and theory-of-mind networks is biased towards seeing a person that had previously been paired with information consistent with group bias (positive for in-group and negative for out-group). This demonstrates how brain regions associated with visual analysis of others and belief reasoning exchange and integrate signals when evaluating in-group and out-group members. The results update models of person perception by showing how and when interplay occurs between perceptual and extended systems when developing a representation of another person.

Introduction

Group biases are prevalent in daily social interactions and typically involve in-group favouritism and out-group dislike (Allport, 1954; Brewer, 1999). To date, neuroscience research has identified a set of brain circuits that control social interactions based on group membership, which span perceptual, affective and cognitive processes (Molenberghs, 2013; Amodio, 2014). However, it is currently unclear how signals from segregated patches of cortex are integrated during the perception of in-group and out-group members. The current fMRI experiment investigates the contribution of functional integration to group bias modulation of person perception.

Among the features used to categorize individuals as members of an in-group or out-group, race is commonly studied (Ito and Bartholow, 2009; Kubota et al., 2012; Azevedo et al., 2013; Molenberghs, 2013). For example, it has been demonstrated that the ability to recognise members of another race is impaired compared to own-race recognition (Malpass and Kravitz, 1969). Besides such pre-existing social categories, group biases can also be elicited by assigning individuals to a group based on arbitrary rules, such as the toss of a coin; a procedure known as minimal group assignment (Tajfel et al., 1971). Such an arbitrary categorisation also leads to better recognition of in-group members (Bernstein et al., 2007), as well as more favourable judgments of in-group compared to out-group members (Tajfel et al., 1971; Otten and Moskowitz, 2000; Hertel and Kerr, 2001). As such, even a temporary group assignment based on arbitrary criteria biases the way others are perceived and judged. In short, group membership has a powerful influence on the mental operations that underpin and guide social interactions.

Over the last 15 years, neuroscience research has started to investigate the neural correlates of group-bias. Consistent with the majority of human cognitive neuroscience research (Fox and Friston, 2012), investigations into the neural correlates of group bias have

primarily focussed on measuring the response of functionally segregated brain circuits. These studies have shown that several brain circuits that span perceptual, affective, and cognitive systems are sensitive to group membership (Figure 1; Molenberghs 2013; Amodio 2014). For example, patches of cortex along the ventral visual stream, which are involved in person perception (Kanwisher, 2010), show a response bias for in-group compared to out-group members based on racial and minimal group assignment (Golby et al., 2001; Van Bavel et al., 2008, 2011; Azevedo et al., 2013). Reduced activity for out-group compared to in-group members has been associated with diminished motivation to individuate out-group members (Malpass and Kravitz, 1969; Golby et al., 2001).

When categorising others we also “feel” differently about in-group compared to out-group members (Harris and Fiske, 2007; Mackie et al., 2008; Azevedo et al., 2013). An “affective network” of brain regions comprising amygdala, insula, striatum, and anterior frontal cortex, has been found to be underpin the ability to feel what someone else might feel (Keysers and Gazzola, 2009). This affective network also shows sensitivity to group biases (Golby et al., 2001; Wheeler and Fiske, 2005; Eres and Molenberghs, 2013; Molenberghs, 2013; Amodio, 2014; Azevedo et al., 2014; Molenberghs et al., 2016). For instance, left OFC was more active when participants saw an out-group member inflict harm to an in-group member compared to an out-group member (Molenberghs et al., 2016). Moreover, this area was functionally coupled with left insula and amygdala under these conditions, revealing a bias in the affective network to preferentially process in-group suffering.

A third neural network to show sensitivity to group membership is the Theory-of-Mind (ToM) network (Harris and Fiske 2007; Volz et al. 2009; Contreras et al. 2012; Eres and Molenberghs 2013; Molenberghs and Morrison 2014). The ToM-network is engaged when making self-other distinctions, when reasoning about others’ mental states (cognitive empathy), as well as when inferring traits about others (van Overwalle, 2009). The ToM-

network includes mPFC, temporal poles, temporoparietal junction (TPJ), and precuneus (Frith and Frith 1999; Mitchell, 2009; Saxe and Kanwisher 2003; van Overwalle 2009). When categorising individuals as in-group members, several ToM nodes are also involved (Volz et al., 2009; Molenberghs and Morrison, 2014). For example, when dividing money between in- and out-group members, participants gave more money to their in-group members and this decision was accompanied by greater activation of mPFC and left TPJ (Volz et al., 2009). Volz and colleagues (2009) suggest that ToM-network engagement reflects the different demands placed on self-other judgments when evaluating in-group compared to out-group members.

In sum, prior neuroimaging studies have shown how segregated patches of cortex are associated with seeing “us” and “them” during social interactions (Molenberghs, 2013). A key question from a neuroscience perspective, however, is how distributed neural circuits interact to support mental processes (Sporns et al., 2005; Sporns, 2014). Indeed, mental processes are likely to be an emergent property of network integration, rather than the sole work of segregated groups of neurons acting alone (Yuste, 2015). Network models of brain function that comprise interacting components have been proposed and supported in theoretical and systems biology (Bassett and Gazzaniga, 2011), but few empirical studies have directly tested how segregated circuits exchange information. For instance, with regard to group bias, it is currently unclear to what extent and in what ways neural circuits interact as a function of group membership. The current fMRI study uses functional connectivity analyses to investigate group bias modulation of person perception.

The design of the study was based on evidence that in-group members are viewed more positively than out-group members (Allport, 1954; Mullen et al., 1992; Brewer, 1999), as well as on research revealing that information consistent with stereotypes is remembered better than bias-inconsistent information (Fyock and Stangor, 1994). We hypothesised

increased functional coupling between perceptual (Fusiform and Extrastriate Body Areas, FBA and EBA), affective, and cognitive (ToM) neural networks when seeing a person that had previously been paired with information consistent with their biases (positive for in-group and negative for out-group). Prior neuroimaging work has shown that body and ToM networks show increased coupling when forming links between body cues and social knowledge (Greven et al., 2016), as well as recalling social knowledge based on body cues (Greven & Ramsey, 2017). As such, the current study would extend prior work by understanding how neural network integration supports group bias modulation of person perception. Although more group bias research in person perception has focussed on faces, bodies convey a multitude of relevant social signals and offer cues that faces might hide (Slaughter et al., 2004; Aviezer et al., 2012), which makes bodies interesting to study in their own right. More generally, as integration between discrete brain circuits is a growing consideration for understanding brain function (Friston and Price, 2001; Sporns et al., 2005; Sporns, 2013) understanding how perceptual, cognitive and affective networks interact is a model problem that speaks to a fundamental question in human neuroscience.

Materials and Methods

Participants

Twenty-four participants (15 females; mean \pm SD age: 22.6 ± 4.7 years) were recruited from the Bangor community and received a monetary reimbursement of £15 for completing the fMRI experiment. All participants had normal or corrected-to-normal vision and reported no history of neurological damage and gave informed consent according to the local ethics guidelines. A behavioural pilot experiment was completed to validate the task and involved 31 participants (24 females; mean \pm SD age: 20.8 ± 6 years). No participants

completed both pilot and fMRI experiments. For 3 participants, 2 sessions from the main task had to be removed due to excessive head motion (displacement above 3 mm).

Overview of the experiment

The full experimental design comprised a 3 (Social knowledge: Positive, Negative, Neutral) x 2 (Group bias: in-group, out-group) factorial design. In order to study group bias modulation of person perception, the current study only analysed Positive and Negative social knowledge conditions. All analyses in the current experiment, therefore, focus on a 2 (Social knowledge: Positive, Negative) x 2 (Group bias: in-group, out-group) factorial design. Analyses investigating the recall of social knowledge compared to neutral knowledge have been reported elsewhere (Greven and Ramsey, 2017).

The experimental paradigm consisted of several phases (Figure 2): 1) Group assignment to the yellow or blue team; 2) Encoding phase, where participants formed an impression of a person based on presentation of a body and a statement; 3) fMRI experiment, where participants observed all the bodies from the encoding phase and were asked to recall knowledge about each person; 4) Recognition phase, where participants had to judge which of the two bodies presented in each trial was previously paired with a statement that was presented. Further details of each phase of the experiment are provided below. Before the scanning experiment, a pilot behavioural experiment was completed to validate the stimuli and ensure that participants could reliably associate and recall information about bodies (Supplementary Methods).

Stimuli

Pictures of 128 bodies were adapted from Greven et al. (2016) that had been selected to convey an emotionally-neutral posture (i.e., crossed-arms or slouching postures were not

included) but varied in terms of body shape, skin colour and clothing. 16 extra pictures (8 female) were added to the 128 pictures from Greven and colleagues (2016) for a total of 144 bodies (72 female). The racial make-up of the bodies was balanced for blue and yellow groups. Consistent with prior work (Downing et al., 2007), in order to target regions selective for images of bodies and not faces, images had been cropped so the head was not visible. The bodies were edited using GIMP 2.8 software (www.gimp.org) to give them a blue and a yellow shirt (each body could be part of either team). Participants would never see the same body in both a yellow and a blue shirt. Instead, half the participants would see bodies 1 – 72 in blue and 73 – 144 in yellow, and the other participants would see the opposite combination. Each body was only shown once during the encoding experiment, to avoid any possible effects of combining the same person with different social knowledge statements over the course of the experiment.

Social knowledge stimuli comprised 144 statements that were adapted from Mitchell et al. (2006) to convey either trait-based (positive and negative) or neutral information. An example of a trait-implying statement is “He cut in front of the man in line”, implying the person is inconsiderate, whereas a neutral example is “She walked through the swivel doors”. Each statement (48 positive, 48 negative, 48 neutral) was presented once during the experiment.

Behavioural tasks

Group assignment: Each participant was assigned to one of two teams upon arrival. They believed this happened randomly as they picked one of two coins (blue or yellow) out of a bag. In fact, it was ensured that there were an equal number of females and males in each team. For this purpose, coins were occasionally both of the same colour, unbeknownst to the participant. After being assigned to a team, participants wore a blue or yellow t-shirt

depending on their group assignment and completed a group association task in order to enforce their association with their team members. In the group association task, participants were presented with every single body they would later see in the fMRI experiment. They had to answer as fast and accurately as possible to which team this person belonged by pressing 'F' for their team and 'J' for the other team.

Encoding phase: In the encoding task participants were told that they would see lots of different bodies about whom they would learn something, and later on they would be asked a number of questions about the bodies. In each trial, participants were presented concurrently with a body (wearing a blue or yellow shirt) and a social knowledge statement which could be positive, negative, or neutral. For each participant, bodies were randomly assigned to the statements. Thus, there was no systematic relationship between particular bodies and statements across participants, which removes any coupling between low-level stimulus artefacts and any one condition in our design.

The body (full-colour picture, 300 x 750 pixels) was presented in the middle of the screen with text underneath (fontsize 30 pt, 250 pixels below the centre of the screen). Each trial started with the presentation of a fixation cross for 500 ms, followed by the simultaneous presentation of an agent and a statement for 5000 ms. Participants were instructed to pay attention to both the person as well as to the knowledge that they would receive about that person.

There were 144 trials in the encoding phase (24 per condition; Positive, Negative, and Neutral for Blue and Yellow teams). Trials were presented in 8 blocks containing a random sequence of 18 trials from 3 valence conditions. Blocks alternated between a presentation of team yellow and team blue. To make sure participants paid attention to all aspects of the stimuli, at the end of each block they were asked a yes/no-question about the previous trial. Within a maximum duration of 5 seconds, yes/no responses were made by pressing the 'F'

and 'J' button, respectively. These questions could be about the agent's gender (was this person a man/woman?), or body (was this person facing forward?), as well as the person knowledge statements (did this person touch an object? did this person have a positive/negative attitude?). To ensure that participants remained alert to all elements of these stimuli, the content of questions could not be predicted.

fMRI scanning: Shortly after the encoding phase (approximately 5 minutes), participants entered the scanner. Here, all the bodies were presented again. Participants were instructed to form an impression of these people based on what they previously learned about them. More details on fMRI scanning procedures are given below.

Recognition phase: After completing scanning, participants performed a recognition task where all the bodies and statements were presented again. In each trial, two bodies appeared on the screen (both of the same gender, one of each team) together with a statement. One of the two bodies was previously paired with that statement. During this task, each body was presented twice, once as the correct and once as the incorrect answer. There were six different conditions: Positive (1), Negative (2), and Neutral (3) where team blue was the correct answer, and Positive (4), Negative (5), and Neutral (6) where team yellow was the correct answer.

Behavioural data analysis

A trial was considered an outlier if the reaction time was below 200 ms, ensuring that participants had taken enough time to read the statement and observe the bodies. This resulted in 0.10% rejected trials in the pilot experiment, and 0.67% rejected trials in the post-scanning recognition task. Participants' performance (percent accurate) on the recognition task was first compared for all conditions against chance-level performance (50%). To do so, 95% confidence intervals (CIs) were calculated for each condition compared to 50% and Cohen's

d_z was calculated as a measure of effect size by dividing the mean difference by the standard deviation of the difference (Cohen, 1992; Lakens, 2013).

In addition, a 2 (Valence: Positive, Negative) x 2 (Group: in-group, out-group) ANOVA compared performance between conditions. We expected an interaction between Valence and Group, whereby recognition was better when in-group members were associated with positive compared to negative social knowledge and vice versa for out-group members. A significant interaction would be followed-up with planned contrasts using 95% CIs, where recognition for positive and negative bodies was compared for in- and out-group members separately.

fMRI experiment:

Each participant's scanning session started with a run of the body-localiser (see details below), followed by two runs of the main task. A further body-localiser run and two runs of the main task then followed. Interspersing the body-localiser between runs of the main task was done to vary the experience for participants and offset boredom. Participants then completed two runs of the ToM-localiser (see details below). The ToM-localiser was always presented after participants had completed the main task, to ensure that participants were not primed towards making trait inferences during the main task. Stimuli were presented using a desktop PC and Matlab software with Psychtoolbox (www.psychtoolbox.org).

Main experimental task: The main task used a block-design to enhance statistical power (Friston et al., 1999) with blocks of bodies presented for 16 seconds. Each image (300 x 650 pixels) was presented for 1800 ms, followed by a blank screen for 200 ms, resulting in a total of 8 bodies per block. The same bodies presented in the encoding task were now presented during scanning and grouped together in a block according to their assigned social knowledge (positive, negative, and neutral). For example, in a 'positive' block, all 8 bodies

were previously associated with positive social information. Participants were given the instruction to form an impression of each body, based on the information they learned about that body during the encoding phase. Participants were not required to remember the exact details of the social knowledge, but rather to recall the impression (good, bad, or neutral) they previously formed. At the end of each block, participants were asked a question about the previous body relating to their gender (was this person a woman/man?) or their team (was this person part of your/other team?). From trial-to-trial, the image location was slightly jittered (4 different locations that varied by 10 pixels around a central fixation dot). From the four options, the location of the image on each trial was randomly selected.

In one functional run, 20 blocks were completed and blocks were separated by a jittered rest block with an average duration of 7 seconds (which varied between 5 and 9 seconds with 500 ms steps). The first 10 blocks in a run showed one team (Blue or Yellow) and the remaining blocks showed the other team. The order of team presentation was counterbalanced both within and across participants. For each participant, the first and last run had a fixed order (e.g., yellow followed by blue), and the second and third runs had the opposite order (e.g., blue followed by yellow). For half of the participants, the first run showed yellow then blue and the other half of participants showed blue followed by yellow. Therefore, each functional run was composed of two 10-block sequences. Each 10-block sequence showed bodies from a single team with each block showing bodies from one condition (Positive, Negative, or Neutral). In order to help effectively model the influence of different events on BOLD signal, block order within each 10-block sequence was counterbalanced so that within each sequence, each condition was preceded equally often by all conditions (Josephs and Henson, 1999; Wager and Nichols, 2003; Aguirre, 2007). To provide a completely balanced block “history” across conditions, each sequence began with a “starter block”, which was not included in the data analysis but modelled as a covariate of no

interest. Subsequently, three further blocks from each Social Knowledge condition were presented in a counterbalanced manner. Each participant completed 4 functional runs of this task, with 24 Positive (half blue, half yellow), 24 Negative and 24 Neutral blocks across the experiment for a total of 96 trials per condition. For all subsequent analyses, we focus on Positive and Negative conditions for In- and Out-group bodies.

Functional localisers: To localise body-selective brain regions we used an established paradigm (Downing et al., 2007; <http://pages.bangor.ac.uk/~pss811/page7/page7.html>). We presented 12-sec blocks of cars and of whole bodies (without heads). A run started with a blank screen for 14 seconds, followed by two alternations of each condition. This was repeated a second time, and followed by a final rest period of 14 seconds. Each image was presented for 600 ms, followed by a blank screen for 100 ms. Twice during each block, the same image was presented two times in a row. Participants had to press a button whenever they detected this immediate repetition (1-back task). The image location was slightly jittered in the same way as in the main task. Each participant completed two runs of this task, counterbalancing the order of the stimulus presentation (Bodies or Cars).

To localise brain regions that respond to mental state reasoning, we used an established ToM-localiser (Dodell-Feder et al., 2011; <http://saxelab.mit.edu/superloc.php>). Participants read 10 short false belief stories, in which the belief characters have about the state of the world is false. Participants also read 10 false photograph stories, where a photograph, map, or sign has out-dated or misleading information. After reading each story, participants had to answer whether the subsequently presented statement is true or false. Each run started with a 12 second rest period, after which the stories and questions were presented for 14 seconds combined (stories: 10 seconds; questions: 4 seconds), and were separated by a 12 second rest period. The order of items and conditions is identical for each subject. In the

first run, stimuli 1 – 5 from each condition were presented, and the remaining stimuli were presented during the second run.

Data Acquisition

The experiment was conducted on a 3 Tesla scanner (Philips Achieva), equipped with a 32-channel SENSE-head coil. Stimuli were displayed on a MR safe BOLD screen (Cambridge Research Systems: <http://www.crs Ltd.com/>) behind the scanner, which participants viewed via a mirror mounted on the head-coil. T2*-weighted functional images were acquired using a gradient echo echo-planar imaging (EPI) sequence. An acquisition time of 2000 ms was used (image resolution: $3.03 \times 3.03 \times 4 \text{ mm}^3$, TE = 30, flip angle = 90°). After the functional runs were completed, a high-resolution T1-weighted structural image was acquired for each participant (voxel size = 1 mm^3 , TE = 3.8 ms, flip angle = 8° , FoV = $288 \times 232 \times 175 \text{ mm}^3$). Four dummy scans ($4 * 2000 \text{ ms}$) were routinely acquired at the start of each functional run and were excluded from analysis.

Data preprocessing and analysis

Data were preprocessed and analysed using SPM8 (Wellcome Trust Department of Cognitive Neurology, London, UK: www.fil.ion.ucl.ac.uk/spm/). Functional images were realigned, unwarped, corrected for slice timing, and normalized to the MNI template with a resolution of $3 \times 3 \times 3 \text{ mm}$ and spatially smoothed using an 8 mm smoothing kernel. Head motion was examined for each functional run and a run was not analysed further if displacement across the scan exceeded 3 millimetres.

Univariate model and analysis: Each condition was modelled from the onset of the first body for a duration of 16 seconds. A design matrix was fitted for each participant with 7

regressors in total; one for each condition (PosIn, PosOut, NegIn, NegOut, NeutralIn, NeutralOut), and one for the starter blocks.

Main effects of Valence [Pos > Neg; Neg > Pos] and Group [In > Out; Out > In] were calculated first for completeness. The main analysis of interest, however, was the Valence by Group interaction [(PosIn > PosOut) > (NegIn > NegOut)] because it tests our primary hypothesis that brain systems will be tuned to information that is consistent with group biases (positive for in-group and negative for out-group). This univariate interaction analysis served two functions. As our primary research question could only be addressed by functional connectivity analyses, the first function of univariate analysis was to identify seed regions for subsequent connectivity-based analyses. The second function enabled the test of magnitude-based hypotheses regarding the role of body perception, affective, and ToM networks during the perception of bodies as a function of group bias. That is, we will be able to test if body, affective and ToM networks are preferentially involved when visually processing bodies about which particular trait and group-based information is known.

For the body and ToM localiser, a design matrix was fitted for each participant with 2 regressors, two for each condition (bodies and cars; false beliefs and false photographs). Body-selective regions were revealed by contrasting bodies and cars (Bodies > Cars). The ToM-network was revealed by contrasting false beliefs with false photographs (False Beliefs > False Photographs).

Psychophysiological Interaction analysis: Our primary hypothesis was that body-selective areas, as well as parts of the affective and ToM networks would interact more when recalling trait information that fits the participant's group bias (positive for in-group members and negative for out-group members). To test this hypothesis, we used psychophysiological interaction (PPI) analysis (Friston et al., 1997). PPI enables the identification of brain regions whose activity correlates with the activity of a seed region as a function of a task. Here we

used a generalized form of PPI, which allows for comparisons across the complete design space (McLaren et al., 2012). By doing so, it is possible to see whether any voxels across the brain show a correlation with activity in the seed region (the “physiological” element) as a function of the two conditions within the main task (the “psychological” element) (Figure 3C).

Two steps were taken to define seed regions for the PPI analysis (Figure 3A). First, based on the group-level univariate analysis, we identified any clusters of overlap between the interaction contrast and the functional localisers (i.e., body and/or ToM localiser) at the group-level. This group-level analysis can identify clusters showing body or ToM selectivity as well as sensitivity to the main task’s contrast. Second, if clusters of overlap were identified at the group-level, we identified subject-specific coordinates for regions of overlap at the single-subject level, thus allowing for inter-individual differences in peak responses. Separately for each individual participant we searched for overlap between the interaction contrast and the functional localisers (body and/or ToM localiser at the single-subject level). In order to include as many participant’s data as possible, we searched for overlap across a range of thresholds ($p < .001 - 0.5$), which is common when identifying seed regions in individual’s data (Spunt and Lieberman, 2012; Klapper et al., 2014; Paulus et al., 2015). For each seed region, therefore, we report how many participants show overlap between the main task’s contrast (across a range of thresholds; reported in Supplementary Table 2) and functional localisers at a fixed threshold ($p < .001$, voxel extent = 10). Volumes were generated using a 6 mm sphere, which was positioned on each individual’s seed-region peak. PPI analyses were run for all seed regions that were identified in this manner.

PPI models for each participant included the 7 regressors from the univariate analyses as covariates of no interest, as well as 8 PPI regressors. PPI regressors included one for each condition (6 in total), one for the starter block, and one that modelled seed region activity

(Figure 3B). The starter block, seed region and neutral condition regressors are also modelled as covariates of no interest. Although we use clusters emerging from the univariate analysis to define seed regions for the PPI analysis, our PPI analysis is not circular (Kriegeskorte et al., 2009), because all regressors from the univariate analysis are included within the PPI model as covariates of no interest (O'Reilly et al., 2012). The PPI analyses, therefore, explain variance in addition to that which is already explained by other regressors in the design and is statistically independent to the univariate analysis.

To create the PPI regressors, the time series in the seed region was specified as the first eigenvariate, and was consequently deconvolved to estimate the underlying neural activity (Gitelman et al., 2003). Then, the deconvolved time series was multiplied by the predicted, pre-convolved time series of each of the seven regressors (6 conditions, and 1 starter block). The resulting PPI for each condition in terms of predicted “neural” activity was then convolved with the canonical haemodynamic response function (HRF) and the time series of the seed region as covariates of no interest (Klapper et al., 2014; McLaren et al., 2012; Spunt and Lieberman, 2012). At the second-level analysis, we examined the same Valence by Group interaction contrast as in the univariate analyses [(PosIn > PosOut) > (NegIn > NegOut)].

For all group-level analyses (univariate and connectivity-based), images were thresholded using a voxel-level threshold of $p < .001$ and a voxel-extent of 10 voxels. Based on our hypotheses for functional connections between core and extended body perception networks, we inclusively mask the contrasts from the main task by body and ToM localisers (Bodies > Cars and False Beliefs > False Photographs thresholded at $p < .001$, $k = 10$). In addition, an affective network mask was created by centring 15 mm spheres on coordinates taken from prior literature, which had identified brain regions associated with affective responses. The affective network mask included amygdala, insula, striatum and orbital frontal cortex (details

of all masks are reported in Table 1). Inclusive masking in this manner makes sure that only responses in brain regions associated with body perception, affective processing and ToM are interpreted. The results from these analyses are presented in Tables 2 and 3. P values following correction for multiple comparisons at the cluster level (Friston et al., 1994) are also reported. To localise functional responses we used the anatomy toolbox (Eickhoff et al., 2005).

Results

Behavioural data

Pilot data. Performance on the recognition task revealed that, after an average retention time of 8.81 +/- 1.96 minutes, participants performed above chance on all conditions (PosIn: M=69.50%, CI.95 [64.27, 74.73], Cohen's $d_z=1.37$; PosOut: M=64.11%, CI.95 [58.59, 69.63], Cohen's $d_z=0.94$; NegIn: M=62.10%, CI.95 [54.36, 69.84], Cohen's $d_z=0.57$; NegOut: M=66.47%, CI.95 [59.55, 73.38], Cohen's $d_z=0.87$). There was no main effect of Valence or Group (Valence: $F(1,30)=1.07$, $p=.31$, $\eta_p^2=.03$; Group: $F(1,30)=0.07$, $p=.79$, $\eta_p^2=.002$), nor a significant Valence*Group interaction ($F(1,30)=1.86$, $p=.18$, $\eta_p^2=.06$).

The results of this pilot study demonstrate that after a short retention period (8-9 minutes), performance on the recognition task was above chance-level for all four conditions and ranged between 62-69% accuracy (Figure 4). In the main experiment, fMRI scanning took place after a similar time period following the encoding phase (5-10 minutes). During scanning, therefore, we were confident that participants would be able to recall positive and negative information about both in-group and out-group members at a rate between 62% and 69%. Although recognition performance shows a trend to be tuned in a manner consistent with group bias (i.e., higher for positive in-group information and negative out-group

information; Figure 4), there was no significant interaction between valence and group. Thus, our pilot data showed relatively weak sensitivity to group-valence pairings.

Post-scanning data. First, we tested if recognition performance was greater than chance for each condition. Performance was above chance (50%) for PosIn ($M=55.69\%$, $CI.95 [50.55, 60.83]$, Cohen's $d_z=0.48$) and NegOut ($M=55.34\%$, $CI.95 [51.23, 59.44]$, Cohen's $d_z=0.56$). However, performance was not different from chance-level for PosOut ($M=49.26\%$, $CI.95 [44.08, 54.43]$, Cohen's $d_z=0.06$) and NegIn ($M=48.64\%$, $CI.95 [43.98, 53.29]$, Cohen's $d_z=0.13$).

Second, we tested how performance on the recognition task varied as a function of Valence and Group using a 2x2 ANOVA. There was no main effect of either Valence (Positive or Negative; $F(1,22)=0.07$, $p=.79$, $\eta_p^2=.003$) or Group (In or Out; $F(1,22)=0.002$, $p=.97$, $\eta_p^2<.001$). There was a significant Valence by Group interaction ($F(1,22)=7.71$, $p=.01$, $\eta_p^2=.26$), which showed better recognition of Positive compared to Negative in-group members, and vice versa for out-group members (Figure 4). Follow-up analyses interrogated the interaction by comparing recognition of positive and negative information for in- and out-group members separately. This revealed a difference for the in-group between positive and negative (Mean difference= 7.05% , $CI.95 [2.22, 11.89]$, Cohen's $d_z=0.63$). There was a weaker difference for out-group (Mean difference= 6.08% , $CI.95 [-1.27, 13.43]$, Cohen's $d_z=0.36$). Therefore, the direction of the difference was consistent with predictions for both the in-group, as well as the out-group, but the effect was stronger for the in-group than the out-group.

Compared to the pilot data, which were collected 8-9 minutes after encoding, the post-scanning data were collected 90 minutes after encoding. As such, we suggest that reduced recall performance during the post-scanning recognition phase most likely reflects deterioration of recall performance over time (Figure 4). In addition, there was a Valence by

Group interaction in the post-scanning data but not the pilot data. Even though the interaction term was only significant in the post-scanning data, the pattern of results was consistent throughout both datasets: recall was higher for positive than negative in-group information as well as higher for negative than positive out-group information. We quantitatively confirmed this pattern of results using meta-analysis (Supplementary Results).

The pattern of results is not consistent with a response bias or general tendency to associate in-group members with positive and out-group members with negative information. If results were consistent with a response bias, we would expect the inconsistent conditions to be below chance performance. That is, we would expect in-group targets to be incorrectly associated with good traits (when they previously associated with negative traits) and out-group targets to be incorrectly identified with negative traits (when they were previously associated with positive traits). We do not find this pattern of recognition data in either of our behavioural experiments, which makes a response bias account of our findings unlikely.

Neuroimaging data

Functional localiser data: Group average MNI coordinates across participants are reported in square brackets. For the Bodies > Cars contrast based on the body-localiser data, clusters were revealed in right EBA for all 24 participants [54,-70,4], and in right FBA for 19 participants [51,-40,-23]. For the False Beliefs > False Photographs contrast based on the ToM-localiser data, clusters were revealed in right TPJ [60,-58,25] for 23 participants, and in left TPJ [-45,-64,28], bilateral temporal poles [(-)51,5,-32], Precuneus [-9,-49,34], and mPFC [6,56,28] for 22 participants.

Univariate analyses

Main effects: No suprathreshold clusters were revealed within either the body, ToM, or affective networks for the main effect of Valence or Group (Pos vs. Neg; In vs. Out).

Interaction: The Valence by Group interaction [(PosIn>PosOut) > (NegIn>NegOut)] revealed one cluster in the Affective network mask, which was located in left insula/putamen (Table 2C; Figure 5). The parameter estimates revealed stronger involvement of this cluster during the perception of positive compared to negative in-group members, and vice versa for out-group members. No suprathreshold clusters emerged when masked by the body or ToM localiser. To explore the null result in body and ToM networks further, we lowered the threshold to $p < .005$, $k=0$ and clusters emerged in right fusiform gyrus, which overlapped with the body-localiser (FBA), and bilateral temporal poles and left TPJ, which overlapped with the ToM-localiser (Supplementary Table 1). At this lower threshold, we do not interpret univariate responses in body or ToM networks, due to the increased likelihood of such clusters being false-positives (Eklund et al., 2016). Instead, we use them to guide the location of seed region specification in subsequent PPI analyses. The selection of seed regions in this manner does not influence the integrity of subsequent functional connectivity analyses because the two types of analysis are statistically independent to each other. As a consequence, if seed regions selected in this manner are not part of a network that integrates information as a function of social information and group bias, then we should expect no functional coupling between body, ToM, and affective nodes in subsequent functional connectivity analyses.

Psychophysiological Interaction analyses: Coordinates of overlap within individual participants were identified in right FBA ($n=16$), temporal poles (left: $n=23$; right: $n=18$), left TPJ ($n=19$) and left insula/putamen ($n=19$). We hypothesized that body-selective areas, as well as parts of the affective and ToM networks, would interact more when recalling trait

information that fits the participant's group bias (positive in-group members and negative out-group members).

For the body-selective and affective network seed regions, no suprathreshold clusters emerged (Table 3.1, 3.3). For ToM seed regions, left TPJ was functionally coupled with right FBA in a manner consistent with our prediction. PPI estimates revealed functional coupling that was stronger for positive compared to negative in-group members and vice versa for out-group members (Figure 6, Table 3.2).

At the primary threshold, functional coupling is restricted to one connection between FBA and TPJ. In order to aid future meta-analyses and avoid under-reporting of null-results (Lieberman and Cunningham, 2009), in supplementary material we have included an exploratory set of PPI analyses at a lower statistical threshold ($p < .005$, $k=10$). At this lower threshold, functional coupling is more widespread and involves interactions between perceptual, affective, and ToM networks (Supplementary Tables 3, 4, 5 & Supplementary Figure 2). We do not place any firm interpretations on these results due to the increased likelihood of reporting false-positives (Eklund et al., 2016). Instead, by reporting these findings in supplementary materials, the data will be useful to others who wish to perform meta-analyses or pursue replicating and extending these results. Furthermore, we avoid the file-drawer problem and provide a less biased estimate of effect sizes, thus providing a platform for a more cumulative science (Rosenthal, 1979).

Discussion

Judgments about other people are often biased by favouritism towards in-group members compared to out-group members (Allport, 1954; Brewer, 1999; Stangor, 2014). Previous neuroscience research has revealed that group biases are underpinned by differential engagement of perceptual, affective, and cognitive neural networks (Molenberghs, 2013;

Amodio, 2014). However, the contribution of neural integration between these networks during group bias modulation has received little attention. In the current study, we report functional connectivity between perceptual and cognitive neural networks that is dependent on the combination of valence information and group membership. Specifically, connections between these social brain circuits are tuned to information that is consistent with group biases (i.e., positive for in-group and negative for out-group). In sum, this study demonstrates how functional integration between neural networks is associated with the detection and categorisation of others into “us” and “them”.

Neural network integration during group bias modulation of person perception

The main outcome from this study is clear evidence that the body and ToM networks interact during group bias modulation of person perception. Brain regions associated with visual analysis of others (Kanwisher, 2010) and mental state reasoning (Frith and Frith, 1999; Saxe and Kanwisher, 2003; van Overwalle, 2009) exchange signals when detecting and evaluating in-group and out-group members. Moreover, the exchange of signals has a specific relationship. Functional links between these circuits are greater for bias consistent (positive for in-group and negative for out-group) compared to bias inconsistent pairings. Thus, functional coupling is not squarely centred on group status (in or out-group) or valence of social knowledge (positive or negative); instead, these networks are tuned to particular combinations of social information (in-group, good; out-group, bad). The results therefore provide neural integration evidence for an information processing account that favours bias-consistent pairings (Fyock and Stangor, 1994).

Functional coupling in the current study was specifically tied to particular nodes within each network: Right FBA was coupled to TPJ. Right FBA forms part of a circuit that performs a visual analysis of physical features and is more sensitive to a holistic or whole-

person body representation (Peelen and Downing, 2007; Downing and Peelen, 2011). The current results show that a signal from right FBA biases processing and/or is modulated by processing in the ToM-network. Prior functional connectivity research has shown that forming links between physical features and trait knowledge is associated with coupling between right FBA and the ToM-network (Greven et al., 2016), whereas recall of trait knowledge during body perception is associated with coupling between right EBA and the ToM network (Greven & Ramsey, 2017). We extend this line of research and add to understanding of the connectivity profile of right FBA. We show that right FBA's functional connections are not restricted to forming impressions during first encounters (Greven et al., 2016), but also operate when person perception cues the recall and assimilation of trait knowledge and group status. These findings suggest a broader role for right FBA in social perception, one that not only processes body shape and posture but also exchanges information with other circuits to inform a more global level of identity representation.

The current results provide support for the view that dynamic interplay exists between perceptual and extended systems (Patterson et al., 2007; Collins and Olson, 2014a), and that such interplay indexes a global representation of identity, which is not restricted to physical features, but incorporates a broader landscape of person factors, such as group status and trait knowledge (Ramsey et al., 2011). Moreover, the results in right FBA support the view that category-selective responses in ventral temporal cortex cannot be reduced to visual processing of object features alone, but instead reflect wider knowledge that is tied to the observed object (Bi et al., 2016; Peelen and Downing, 2017).

Toward a network model of group bias

Several distinct mental operations have been shown to underpin the detection and categorisation of others into in-groups and out-groups, which span perceptual, affective and

cognitive processing components (Molenberghs, 2013; Amodio, 2014). We extend these models of group bias by placing them within a network model framework, which considers integration across processing components, as well as local processing within components (Sporns, 2013). As such, we have started to place greater emphasis on the contribution of neural integration to group bias and person perception research (Supplementary Figure 2). From a network model perspective, many questions remain unanswered (Bassett and Gazzaniga, 2011; Wig et al., 2011). Network models have particular structures, which include ‘hubs’ that act as a conduit for information flow between distinct processing components. Future work in social perception may consider investigating whether particular nodes within social circuits act as hubs. Good candidates may be right FBA and temporal poles, due to their anatomical connectivity to each other (Collins and Olson, 2014b) and functional properties (Downing and Peelen, 2011; Olson et al., 2013; Wang et al., 2017), but future work would have to investigate this directly.

Limitations and future directions

In the current study, our results did not survive correction for multiple comparisons and therefore could reflect a Type-1 error or false positive (Eklund et al., 2016). It should be noted, however, that we used functional and anatomical masks to constrain our search space only to three brain networks that were identified based upon prior research investigating the neuroscience of group bias (Molenberghs 2013; Amodio 2014). Therefore, there was a strong *a priori* justification to expect interactions between these specific brain networks and based on our data our best estimate is that the targeted effects may be relatively small in magnitude. Following recent suggestions in the neuroimaging literature (Lieberman and Cunningham, 2009), as well as more generally in psychological science (Cumming, 2014), accurate estimation of population effect sizes will only be possible if all results are reported

transparently even if they are null or mixed results (Simmons et al., 2011; Cumming, 2014). Thus, we offer caution in interpreting our results as they could reflect a false positive, whilst encouraging future studies to test the general hypothesis further as well as consider meta-analytical approaches.

The measure of connectivity used in the current study is correlational and based on functional activity. Therefore, the current study provides no insight into the direction of influence or underlying neural pathway that controls functional coupling between brain areas. Future work using directional connectivity measures (Friston, 2009), structural connectivity (Le Bihan, 2012) and studies that combine fMRI with neurostimulation techniques (Driver et al., 2010; Bestmann and Feredoes, 2013) would provide grounds for more causal inferences to be made regarding network interactions.

A further consideration is that post-scanning recognition performance (approximately 90 min after encoding) is at chance-level for two conditions (inconsistent pairings), but remains above-chance for two conditions (consistent pairings; Figure 4). As such, it is possible that our neural results reflect a more general effect associated with recall of information compared to chance-level recall, rather than any relationship between group bias and social knowledge pairings. However, our pilot data makes this possibility unlikely. Participants were scanned approximately 5-10 minutes after the encoding phase and the main task was being performed for approximately 1 hour in the scanner. We know from our pilot data that at the start of scanning recognition accuracy was between 62-70%. As such, although recall performance decreased over 90 minutes, based on our pilot data we expect that for the majority of scanning, participants' recall was better than chance for all conditions. Consequently, we expect our results to reflect the influence of group bias, even if they underestimate the true size of the effects. Future studies that use a stronger group bias manipulation would be able to test this proposal.

Acknowledgements

This work was funded by a grant from the Economic and Social Research Council (grant number: ES/K001884/1 to R.R.).

References

- Aguirre GK (2007) Continuous carry-over designs for fMRI. *NeuroImage* 35:1480–1494.
- Allport GW (1954) *The Nature of Prejudice*. Cambridge, Massachusetts: Addison-Wesley Publication company.
- Amodio DM (2014) The neuroscience of prejudice and stereotyping. *Nature Reviews Neuroscience* 15:670–682.
- Aviezer H, Trope Y, Todorov A (2012) Body cues, not facial expressions, discriminate between intense positive and negative emotions. *Science* 338:1225–1229.
- Azevedo RT, Macaluso E, Avenanti A, Santangelo V, Cazzato V, Aglioti SM (2013) Their pain is not our pain: Brain and autonomic correlates of empathic resonance with the pain of same and different race individuals. *Human Brain Mapping* 34:3168–3181.
- Azevedo RT, Macaluso E, Viola V, Sani G, Aglioti SM (2014) Weighing the stigma of weight: An fMRI study of neural reactivity to the pain of obese individuals. *NeuroImage* 91:109–119.
- Ball T, Derix J, Wentlandt J, Wieckhorst B, Speck O, Schulze-Bonhage A, Mutschler I (2009) Anatomical specificity of functional amygdala imaging of responses to stimuli with positive and negative emotional valence. *Journal of Neuroscience Methods* 180:57–70.
- Bartra O, McGuire JT, Kable JW (2013) The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *NeuroImage* 76:412–427.
- Bassett DS, Gazzaniga MS (2011) Understanding complexity in the human brain. *Trends in Cognitive Sciences* 15:200–209.
- Bernstein MJ, Young SG, Hugenberg K (2007) The Cross-Category Effect: Mere Social Categorization Is Sufficient to Elicit an Own-Group Bias in Face Recognition. *Psychological Science* 18:706–712.
- Bestmann S, Feredoes E (2013) Combined neurostimulation and neuroimaging in cognitive neuroscience: Past, present, and future. *Annals of the New York Academy of Sciences* 1296:11–30.
- Bi Y, Wang X, Caramazza A (2016) Object Domain and Modality in the Ventral Visual Pathway. *Trends in Cognitive Sciences* 20:282–290.
- Brewer MB (1999) The psychology of prejudice: Ingroup love our outgroup hate? *Journal of Social Issues* 55:429–444.

- Cerliani L, Thomas RM, Jbabdi S, Siero JCW, Nanetti L, Crippa A, Gazzola V, D'Arceuil H, Keysers C (2012) Probabilistic tractography recovers a rostrocaudal trajectory of connectivity variability in the human insular cortex. *Human Brain Mapping* 33:2005–2034.
- Cohen J (1992) A power primer. *Psychological Bulletin* 112:155–159.
- Collins JA, Olson IR (2014a) Knowledge is power: how conceptual knowledge transforms visual cognition. *Psychonomic Bulletin and Review* 21:843–860.
- Collins JA, Olson IR (2014b) Beyond the FFA: the role of the ventral anterior temporal lobes in face processing. *Neuropsychologia* 61:65–79.
- Contreras JM, Banaji MR, Mitchell JP (2012) Dissociable neural correlates of stereotypes and other forms of semantic knowledge. *Social Cognitive and Affective Neuroscience* 7:764–770.
- Cumming G (2014) The New Statistics: Why and How. *Psychological Science* 25:7–29.
- Dodell-Feder D, Koster-Hale J, Bedny M, Saxe RR (2011) fMRI item analysis in a theory of mind task. *NeuroImage* 55:705–712.
- Downing PE, Peelen MV (2011) The role of occipitotemporal body-selective regions in person perception. *Cognitive Neuroscience* 2:186–226.
- Downing PE, Wiggett AJ, Peelen MV (2007) Functional magnetic resonance imaging investigation of overlapping lateral occipitotemporal activations using multi-voxel pattern analysis. *Journal of Neuroscience* 27:226–233.
- Driver J, Blankenburg F, Bestmann S, Ruff CC (2010) New approaches to the study of human brain networks underlying spatial attention and related processes. *Experimental Brain Research* 206:153–162.
- Eickhoff SB, Stephan KE, Mohlberg H, Grefkes C, Fink GR, Amunts K, Zilles K (2005) A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage* 25:1325–1335.
- Eklund A, Nichols TE, Knutsson H (2016) Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences of the United States of America* 113:7900–7905.
- Eres R, Molenberghs P (2013) The influence of group membership on the neural correlates involved in empathy. *Frontiers in Human Neuroscience* 7:1–6.
- Fox PT, Friston KJ (2012) Distributed processing; distributed functions? *NeuroImage* 61:407–426.

- Friston KJ (1994) Functional and Effective Connectivity in Neuroimaging: A Synthesis. *Human Brain Mapping* 2:56–78.
- Friston KJ (2009) Causal modelling and brain connectivity in functional magnetic resonance imaging. *PLoS biology* 7:0220–0225.
- Friston KJ, Buechel C, Fink GR, Morris J, Rolls E, Dolan RJ (1997) Psychophysiological and modulatory interactions in neuroimaging. *NeuroImage* 6:218–229.
- Friston KJ, Price CJ (2001) Generative models, ~ brain function and neuroimaging. *Scandinavian Journal of Psychology* 42:167–177.
- Friston KJ, Zarahn E, Josephs O, Henson RNA, Dale AM (1999) Stochastic Designs in Event-Related fMRI. *NeuroImage* 10:607–619.
- Frith CD, Frith U (1999) Interacting Minds--A Biological Basis. *Science* 286:1692–1695.
- Fyock J, Stangor C (1994) The role of memory biases in stereotype maintenance. *British Journal of Social Psychology* 33:331–343.
- Gitelman DR, Penny WD, Ashburner J, Friston KJ (2003) Modeling regional and psychophysiologic interactions in fMRI: the importance of hemodynamic deconvolution. *NeuroImage* 19:200–207.
- Golby AJ, Gabrieli JDE, Chiao JY, Eberhardt JL (2001) Differential responses in the fusiform region to same-race and other-race faces. *Nature Neuroscience* 4:845–850.
- Greven IM, Downing PE, Ramsey R (2016) Linking person perception to person knowledge in the human brain. *Social Cognitive and Affective Neuroscience* 11:641–651.
- Greven IM, Ramsey R (2017) Person perception involves functional integration between the extrastriate body area and temporal pole. *Neuropsychologia*.
- Harris LT, Fiske ST (2007) Social groups that elicit disgust are differentially processed in mPFC. *Social Cognitive and Affective Neuroscience* 2:45–51.
- Hertel G, Kerr NL (2001) Priming In-Group Favoritism: The Impact of Normative Scripts in the Minimal Group Paradigm. *Journal of Experimental Social Psychology* 37:316–324.
- Ito TA, Bartholow BD (2009) The Neural Correlates of Race. *Trends in Cognitive Sciences* 13:524–531.
- Jakab A, Molnár PP, Bogner P, Béres M, Berényi EL (2012) Connectivity-based parcellation reveals interhemispheric differences in the insula. *Brain Topography* 25:264–271.
- Josephs O, Henson RNA (1999) Event-related functional magnetic resonance imaging: modelling, inference and optimization. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* 354:1215–1228.

- Kahnt T, Chang LJ, Park SQ, Heinze J, Haynes J-D (2012) Connectivity-Based Parcellation of the Human Orbitofrontal Cortex. *Journal of Neuroscience* 32:6240–6250.
- Kanwisher N (2010) Functional specificity in the human brain: a window into the functional architecture of the mind. *Proceedings of the National Academy of Sciences of the United States of America* 107:11163–11170.
- Keysers C, Gazzola V (2009) Expanding the mirror: vicarious activity for actions, emotions, and sensations. *Current Opinion in Neurobiology* 19:666–671.
- Klapper A, Ramsey R, Wigboldus DHJ, Cross ES (2014) The Control of Automatic Imitation Based on Bottom-Up and Top-Down Cues to Animacy: Insights from Brain and Behavior. *Journal of Cognitive Neuroscience* 26:2503–2513.
- Kriegeskorte N, Simmons WK, Bellgowan PSF, Baker CI (2009) Circular analysis in systems neuroscience: the dangers of double dipping. *Nature Neuroscience* 12:535–540.
- Kubota JT, Banaji MR, Phelps EA (2012) The neuroscience of race. *Nature Neuroscience* 15:940–948.
- Kurth F, Zilles K, Fox PT, Laird AR, Eickhoff SB (2010) A link between the systems: functional differentiation and integration within the human insula revealed by meta-analysis. *Brain Structure and Function* 214:519–534.
- Lakens D (2013) Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology* 4:1–12.
- Le Bihan D (2012) Diffusion, confusion and functional MRI. *NeuroImage* 62:1131–1136.
- Lieberman MD, Cunningham WA (2009) Type I and Type II error concerns in fMRI research: re-balancing the scale. *Social Cognitive and Affective Neuroscience* 4:423–428.
- Liu H, Qin W, Qi H, Jiang T, Yu C (2015) Parcellation of the human orbitofrontal cortex based on gray matter volume covariance. *Human Brain Mapping* 36:538–548.
- Mackie DM, Smith ER, Ray DG (2008) Intergroup Emotions and Intergroup Relations. *Social and Personality Psychology Compass* 5:1866–1880.
- Malpass RS, Kravitz J (1969) Recognition for faces of own and other race. *Journal of Personality and Social Psychology* 13:330–334.
- McLaren DG, Ries ML, Xu G, Johnson SC (2012) A generalized form of context-dependent psychophysiological interactions (gPPI): a comparison to standard approaches. *NeuroImage* 61:1277–1286.
- Mitchell JP, Cloutier J, Banaji MR, Macrae CN (2006) Medial prefrontal dissociations during

- processing of trait diagnostic and nondiagnostic person information. *Social Cognitive and Affective Neuroscience* 1:49–55.
- Molenberghs P (2013) The neuroscience of in-group bias. *Neuroscience and Biobehavioral Reviews* 37:1530–1536.
- Molenberghs P, Gapp J, Wang B, Louis WR, Decety J (2016) Increased Moral Sensitivity for Outgroup Perpetrators Harming Ingroup Members. *Cerebral Cortex* 26:225–233.
- Molenberghs P, Morrison S (2014) The role of the medial prefrontal cortex in social categorization. *Social Cognitive and Affective Neuroscience* 9:292–296.
- Mullen B, Brown R, Smith C (1992) Ingroup bias as a function of salience, relevance, and status: An integration. *European Journal of Social Psychology* 22:103–122.
- O'Reilly JX, Woolrich MW, Behrens TEJ, Smith SM, Johansen-Berg H (2012) Tools of the trade: Psychophysiological interactions and functional connectivity. *Social Cognitive and Affective Neuroscience* 7:604–609.
- Olson IR, McCoy D, Klobusicky E, Ross LA (2013) Social cognition and the anterior temporal lobes: a review and theoretical framework. *Social Cognitive and Affective Neuroscience* 8:123–133.
- Otten S, Moskowitz GB (2000) Evidence for Implicit Evaluative In-Group Bias: Affect-Biased Spontaneous Trait Inference in a Minimal Group Paradigm. *Journal of Experimental Social Psychology* 36:77–89.
- Patterson K, Nestor PJ, Rogers TT (2007) Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience* 8:976–987.
- Paulus FM, Müller-Pinzler L, Jansen A, Gazzola V, Krach S (2015) Mentalizing and the Role of the Posterior Superior Temporal Sulcus in Sharing Others' Embarrassment. *Cerebral Cortex* 25:2065–2075.
- Peelen MV, Downing PE (2007) The neural basis of visual body perception. *Nature Reviews Neuroscience* 8:636–648.
- Peelen M V, Downing PE (2017) Category selectivity in human visual cortex: Beyond visual object recognition. *Neuropsychologia*.
- Ramsey R, van Schie HT, Cross ES (2011) No two are the same: Body shape is part of identifying others. *Cognitive Neuroscience* 2:207–208.
- Rosenthal R (1979) The “File Drawer Problem” and Tolerance for Null Results. *Psychological Bulletin* 86:638–641.

- Saxe RR, Kanwisher N (2003) People thinking about thinking people The role of the temporo-parietal junction in “theory of mind.” *NeuroImage* 19:1835–1842.
- Seymour B, Daw N, Dayan P, Singer T, Dolan RJ (2007) Differential Encoding of Losses and Gains in the Human Striatum. *Journal of Neuroscience* 27:4826–4831.
- Simmons JP, Nelson LD, Simonsohn U (2011) False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science* 22:1359–1366.
- Slaughter V, Stone VE, Reed C (2004) Perception of Faces and Bodies: Similar or Different? *Current Directions in Psychological Science* 13:219–223.
- Sporns O (2013) The human connectome: Origins and challenges. *NeuroImage* 80:53–61.
- Sporns O (2014) Contributions and challenges for network models in cognitive neuroscience. *Nature Neuroscience* 17:652–660.
- Sporns O, Tononi G, Kötter R (2005) The human connectome: A structural description of the human brain. *PLoS Computational Biology* 1:0245–0251.
- Spunt RP, Lieberman MD (2012) Dissociating Modality-Specific and Supramodal Neural Systems for Action Understanding. *Journal of Neuroscience* 32:3575–3583.
- Stangor C (2014) *Principles of Social Psychology – 1st International Edition* (Jhangiani R, Tarry H, eds).
- Tajfel H, Billig MG, Bundy RP, Flament C (1971) Social Categorization and Intergroup Behaviour. *European Journal of Social Psychology* 1:149–177.
- Tanaka SC, Doya K, Okada G, Ueda K, Okamoto Y, Yamawaki S (2004) Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. *Nature Neuroscience* 7:887–893.
- Van Bavel JJ, Packer DJ, Cunningham WA (2008) The Neural Substrates of In-Group Bias: A functional magnetic resonance imaging investigation. *Psychological Science* 19:1131–1139.
- Van Bavel JJ, Packer DJ, Cunningham WA (2011) Modulation of the Fusiform Face Area following Minimal Exposure to Motivationally Relevant Faces: Evidence of In-group Enhancement (Not Out-group Disregard). *Journal of Cognitive Neuroscience* 23:3343–3354.
- van Overwalle F (2009) Social cognition and the brain: a meta-analysis. *Human Brain Mapping* 30:829–858.
- Volz KG, Kessler T, von Cramon DY (2009) In-group as part of the self: In-group favoritism

- is mediated by medial prefrontal cortex activation. *Social Neuroscience* 4:244–260.
- Wager TD, Nichols TE (2003) Optimization of experimental design in fMRI: a general framework using a genetic algorithm. *NeuroImage* 18:293–309.
- Wang Y, Collins JA, Koski J, Nugiel T, Metoki A, Olson IR (2017) Dynamic neural architecture for social knowledge retrieval. *Proceedings of the National Academy of Sciences* 114:E3305–E3314.
- Wheeler ME, Fiske ST (2005) Controlling Racial Prejudice. Social-Cognitive Goals Affect Amygdala and Stereotype Activation. *Psychological Science* 16:56–63.
- Wig GS, Schlaggar BL, Petersen SE (2011) Concepts and principles in the analysis of brain networks. *Annals of the New York Academy of Sciences* 1224:126–146.
- Yuste R (2015) From the neuron doctrine to neural networks. *Nature Reviews Neuroscience* 16:487–497.

Table 1. Details of the body, ToM, and affective masks. Average coordinates given for each region of the body-localiser (bilateral extrastriate and fusiform body area; EBA and FBA) and ToM-localiser (bilateral temporoparietal junction (TPJ), temporal poles (TP), Precuneus, and medial Prefrontal Cortex (mPFC)). For the affective mask, coordinates of each area (amygdala, anterior and posterior insula, striatum, and five clusters within the orbitofrontal cortex (OFC)) and their source from the literature are provided.

Network	Area	Hemisphere	Coordinate	Source	
Body	EBA	-	54,-70,4	Functional localiser (Downing et al., 2007)	
	FBA	-	51,-40,-23		
ToM	TPJ	Left	-45,-64,28	Functional localiser (Dodell-Feder et al., 2011)	
		Right	60,-58,25		
	TP	Left	-51,5,-32		
		Right	51,5,-32		
	mPFC	-	6,56,28		
	Precuneus	-	-9,-49,34		
Affective	Amygdala	Left	-20,-3,-20	(Ball et al., 2009)	
		Right	20,-3,-20		
	Anterior insula	Left	-35,12,-4	(Kurth et al., 2010; Cerliani et al., 2012; Jakab et al., 2012; Bartra et al., 2013)	
		Right	37,11,-4		
	Posterior	Left	-38,-9,4		
	insula	Right	39,-6,4		
	Striatum	Left	-16,4,-4		(Tanaka et al., 2004; Seymour et al., 2007; Bartra et al., 2013)
		Right	12,6,4		

Anterior	Left	-6,40,-16	(Kahnt et al., 2012; Liu et al., 2015)
OFC	Right	6,41,-13	
Medial OFC	Left	-16,58,-10	
	Right	14,57,-11	
Posterior	Left	-28,39,-15	
OFC	Right	28,41,-16	
Intermediate	Left	-42,33,-10	
OFC	Right	43,35,-10	
Lateral OFC	Left	-15,23,-21	
	Right	18,23,-21	

Table 2. Univariate results for the Valence by Group [(PosIn > PosOut) > (NegIn > NegOut)] contrast a) masked by the body-localiser, b) masked by the ToM-localiser, and c) masked by the affective network.

Region	Number of voxels	<i>T</i>	<i>p value</i> <i>FWE</i> <i>corrected</i>	Montreal Neurological Institute coordinates		
				x	y	z
<i>a) Masked by body-localiser (EBA and FBA)</i>						
No suprathreshold clusters						
<i>b) Masked by ToM-localiser</i>						
No suprathreshold clusters						
<i>c) Masked by Affective Network</i>						
Left insula	17	3.96	0.20	-30	-16	-2

Table 3. PPI results based on body-selective, theory of mind, and affective network seed regions. Clusters revealed in the PsychoPhysiological Interaction (PPI) analysis for the Valence by Group [(PosIn > PosOut) > (NegIn > NegOut)] contrast using the 1) body selective 2) Theory of mind and 3) Affective network seed regions. Seed regions were defined by the univariate Valence by Group contrast and masked by the Body localiser, ToM-localiser or by the affective network mask.

Region	Number of voxels	<i>T</i>	<i>p value</i> <i>FWE</i> <i>corrected</i>	Montreal Neurological Institute coordinates		
				x	y	z
1) Body-selective seed regions: right FBA						
a) Masked by ToM-localiser						
No suprathreshold clusters						
b) Masked by Affective Network						
No suprathreshold clusters						
2) Theory of mind seed regions						
a) Masked by body-localiser (EBA and FBA)						
Seed regions: bilateral TPs						
No suprathreshold clusters						
Seed regions: left TPJ						
Right fusiform gyrus (FBA)	10	4.31	.25	48	-43	-14
b) Masked by Affective Network						
Seed regions: bilateral TPs and left TPJ						
No suprathreshold clusters						

3) Affective network seed regions: left insula

a) Masked by body-localiser (EBA and FBA)

No suprathreshold clusters

b) Masked by ToM-localiser

No suprathreshold clusters

Supplementary Table 1. Univariate results for the Valence by Group [(PosIn > PosOut) > (NegIn > NegOut)] contrast a) masked by the body-localiser, b) masked by the ToM-localiser.

Region	Number of voxels	<i>T</i>	<i>p value FWE</i> <i>corrected</i>	Montreal Neurological Institute coordinates		
				x	y	z
<i>A) Masked by body-localiser (EBA and FBA)</i>						
<i>Thresholded at p<.005, k=0</i>						
Right fusiform gyrus	5	3.19	.97	45	-37	-14
<i>B) Masked by ToM-localiser</i>						
<i>Thresholded at p<.005, k=0</i>						
Left temporal pole	7	3.69	.83	-42	8	-41
Left temporal pole	2	3.42	.91	-42	23	-20
Right temporal pole	2	3.42	.91	30	17	-29
Left temporal pole	3	2.90	.90	-30	14	-29
Right temporal pole	5	2.99	.86	36	14	-38
Left middle temporal gyrus / temporoparietal junction	1	2.97	.93	-63	-55	16
Right temporal pole	2	2.90	.91	39	17	-23

Supplementary Table 2. Details of individual subjects' overlap between social Valence by Group [(PosIn > PosOut) > (NegIn > NegOut)] contrast and the body and ToM localisers as well as the affective network mask.

Seed-region	<i>Main task threshold at which overlap was found in individual subjects</i>								
	p<.001	p<.005	p<.01	p<.05	p<.1	p<.2	p<.3	p<.4	p<.5
Right FBA (n=16)	1	2	0	2	3	4	3	1	0
Left TPJ (n=19)	2	3	2	2	2	4	2	1	1
Left TP (n=23)	0	0	0	5	6	5	5	1	1
Right TP (n=18)	0	1	0	4	5	4	3	1	0
Left insula (n=19)	0	0	1	6	5	1	3	2	1

Supplementary Table 3. PPI results based on body-selective seed regions. Clusters revealed in the PsychoPhysiological Interaction (PPI) analysis for the Valence by Group [(PosIn > PosOut) > (NegIn > NegOut)] contrast using the body selective seed region defined by the univariate Valence by Group contrast (right FBA), a) masked by the ToM-localiser, and b) masked by the affective network.

Region	Number of voxels	<i>T</i>	<i>p value FWE corrected</i>	Montreal Neurological Institute coordinates		
				x	y	z
<i>a) Seed region: right FBA masked by ToM-localiser</i>						
Right TPJ	13	4.23	.74	63	-46	37
<i>b) Seed region: right FBA masked by affective network</i>						
Left anterior insula	20	4.17	.65	-36	23	4
Left striatum/superior orbital gyrus	18	4.08	.68	-18	17	-14
Left hippocampus extending into amygdala	25	4.07	.57	-15	5	-26
Right middle orbital gyrus	22	3.66	.62	30	38	-14
		3.49		33	29	-17
Right amygdala	10	3.16	.81	15	11	-23
		3.06		18	2	-17

Supplementary Table 4. PPI results based on theory-of-mind seed regions. Clusters revealed in the PsychoPhysiological Interaction (PPI) analysis for the Valence by Group [(PosIn > PosOut) > (NegIn > NegOut)] contrast using ToM seed regions defined by the univariate Valence by Group contrast (bilateral temporal poles (TP) and left TPJ), a) masked by the body-localiser, and b) masked by the affective network.

Region	Number of voxels	<i>T</i>	<i>p value FWE corrected</i>	Montreal Neurological Institute coordinates		
				x	y	z
<i>a) Masked by body-localiser (EBA and FBA)</i>						
<i>Seed regions: bilateral TP</i>						
No suprathreshold clusters						
<i>Seed region: left TPJ</i>						
Right fusiform gyrus (FBA)	70	4.31	.33	48	-43	-14
<i>b) Masked by affective network</i>						
<i>Seed region: right TP</i>						
Right amygdala	10	3.92	.77	18	5	-14
Left amygdala extending into hippocampus	12	3.33	.74	-24	-4	-14
		3.14		-15	-7	-14
<i>Seed region: left TP</i>						
Left insula	31	3.50	.52	-33	8	4
<i>Seed region: left TPJ</i>						
No suprathreshold clusters						

Supplementary Table 5. PPI results based on affective seed regions. Clusters revealed in the PsychoPhysiological Interaction (PPI) analysis for the Valence by Group [(PosIn > PosOut) > (NegIn > NegOut)] contrast using affective seed regions defined by the univariate Valence by Group contrast (left insula), a) masked by the body-localiser, and b) masked by the ToM-localiser.

Region	Number of voxels	<i>T</i>	<i>p value FWE</i> <i>corrected</i>	Montreal Neurological Institute coordinates		
				x	y	z
<i>Affective network seed region: left insula</i>						
<i>a) Masked by body-localiser (EBA and FBA)</i>						
No suprathreshold clusters						
<i>b) Masked by ToM-localiser</i>						
Left anterior superior temporal sulcus	10	3.96	.78	-57	-16	-5
Left anterior superior temporal sulcus	13	3.55	.74	-66	-34	1

Figure 1. Neural networks involved in body perception (green), Theory of Mind (blue), and affective processing (yellow). Abbreviations: Extrastriate Body Area (EBA), Fusiform Body Area (FBA), TemporoParietal Junction (TPJ), Temporal Pole (TP), Precuneus (PreC), medial PreFrontal Cortex (mPFC), Amygdala (AMG), OrbitoFrontal Cortex (OFC).

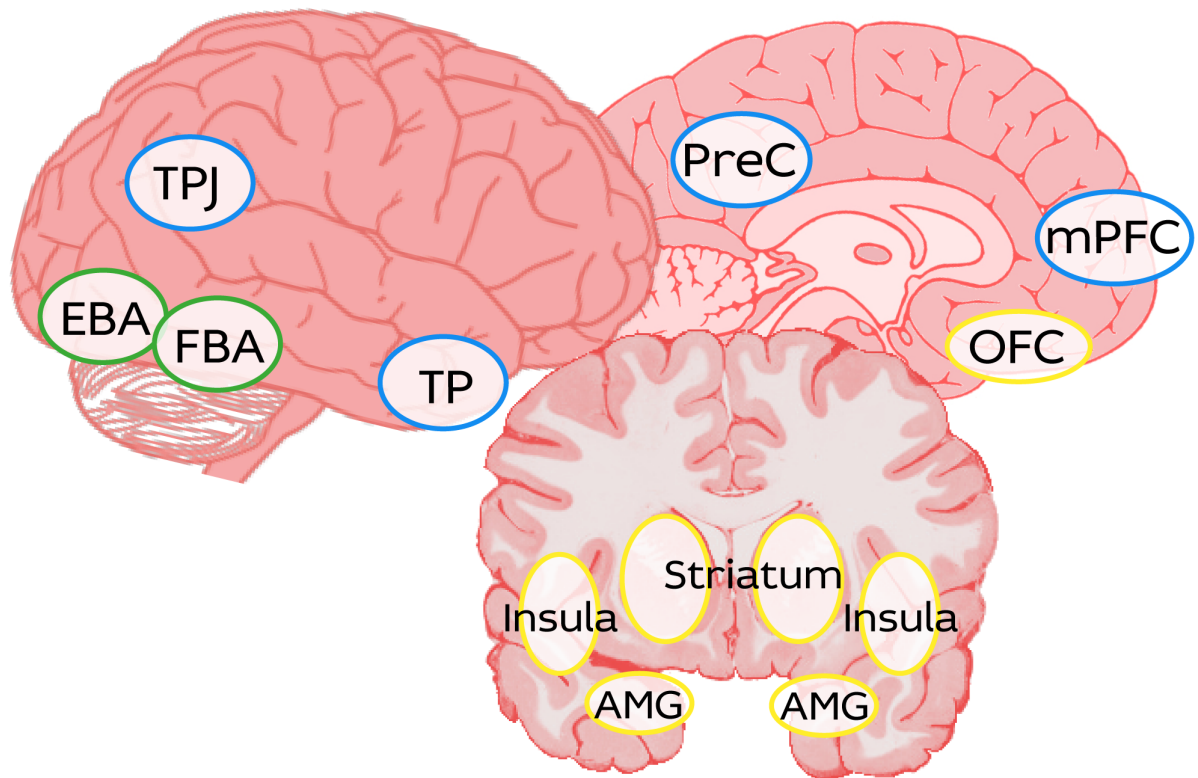


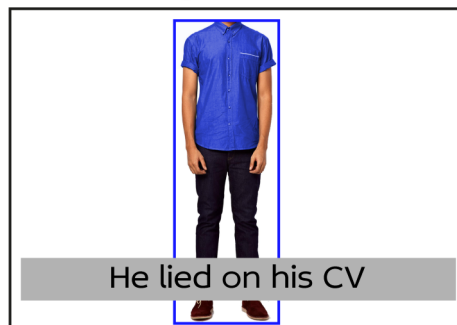
Figure 2. Methods and procedure for the pilot and fMRI experiment.

participant is randomly
assigned to team and gets
t-shirt to wear (blue or yellow)

group association task (~5 min):



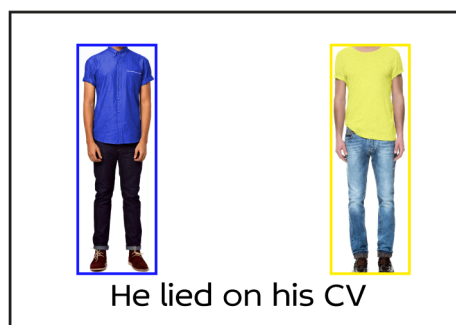
encoding task (~15 min):



fMRI task (~ 1 hour):



recognition task (~7 min):



total duration:
~ 2 hours

Figure 3. Flow chart illustrating the steps to define seed regions and run PsychoPhysiological Interactions (PPI) analyses. A) Identification of seed regions in the univariate analysis was done at group and single-subject level to allow for inter-individual differences in peak responses. B) An illustration of the design matrix (this was the same for each run), that was created for each participant. C) The “psychological” (task) and “physiological” (time course from seed region) inputs for the PPI analysis.

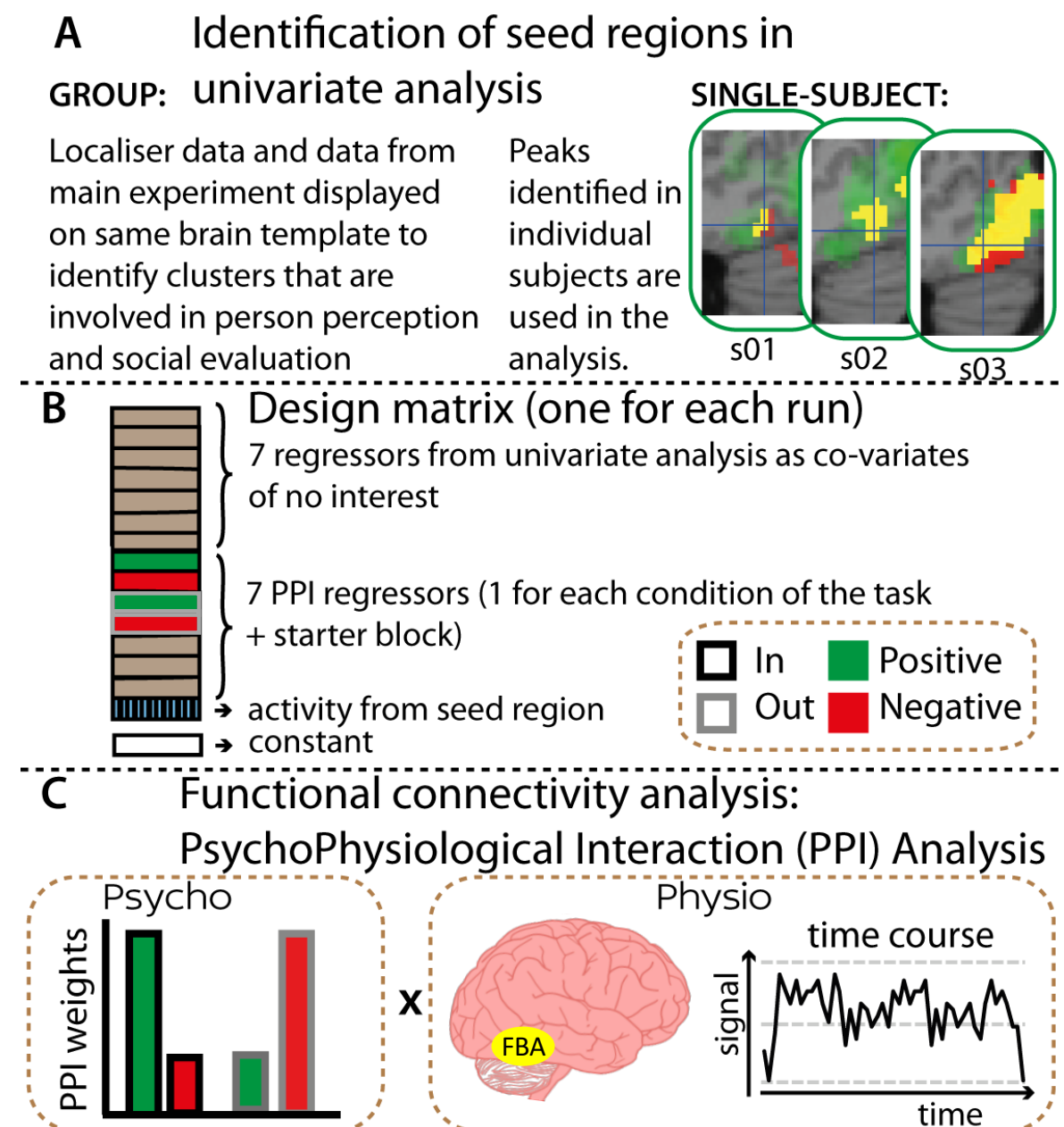


Figure 4. Behavioural results for the pilot and post-scanning recognition task. **: $p < .01$.

Error bars show 95% confidence intervals. The data in the pilot experiment were collected after a 8-9 minute delay, while the post-fMRI data was collected 90 minutes after finishing the encoding task.

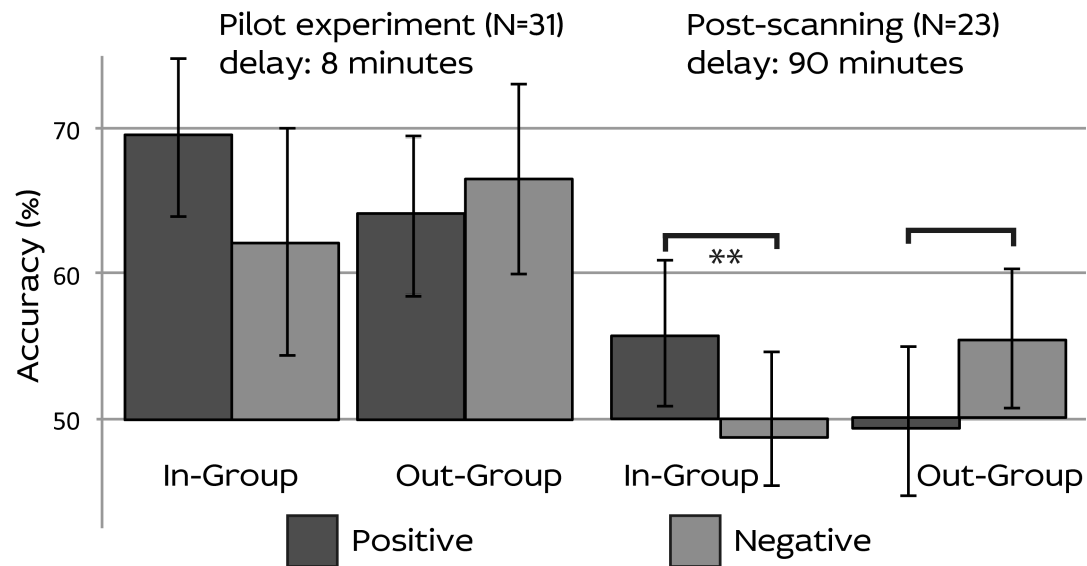


Figure 5. Results from the univariate analysis. The Valence by Group [(PosIn > PosOut) > (NegIn > NegOut)] contrast revealed one cluster within the affective network in left posterior insula. The parameter estimates are extracted from a 4 mm sphere around the peak coordinate.

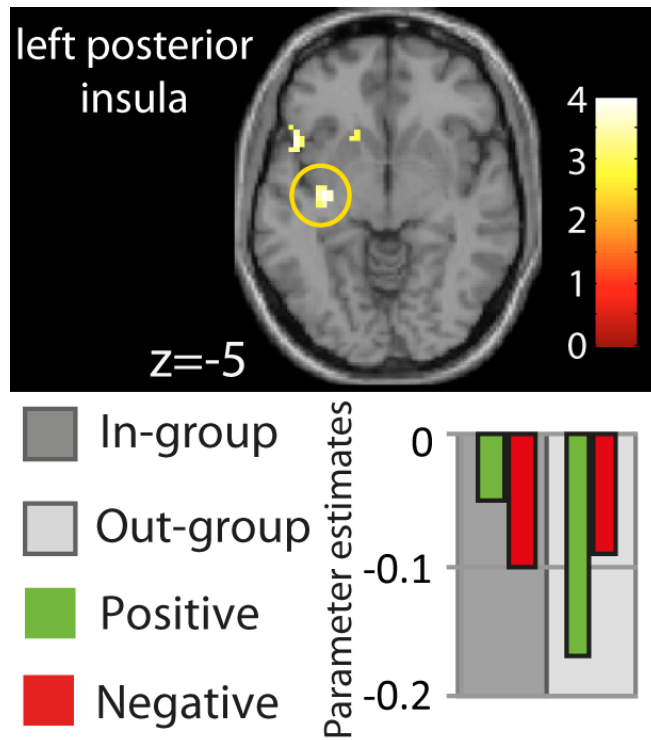
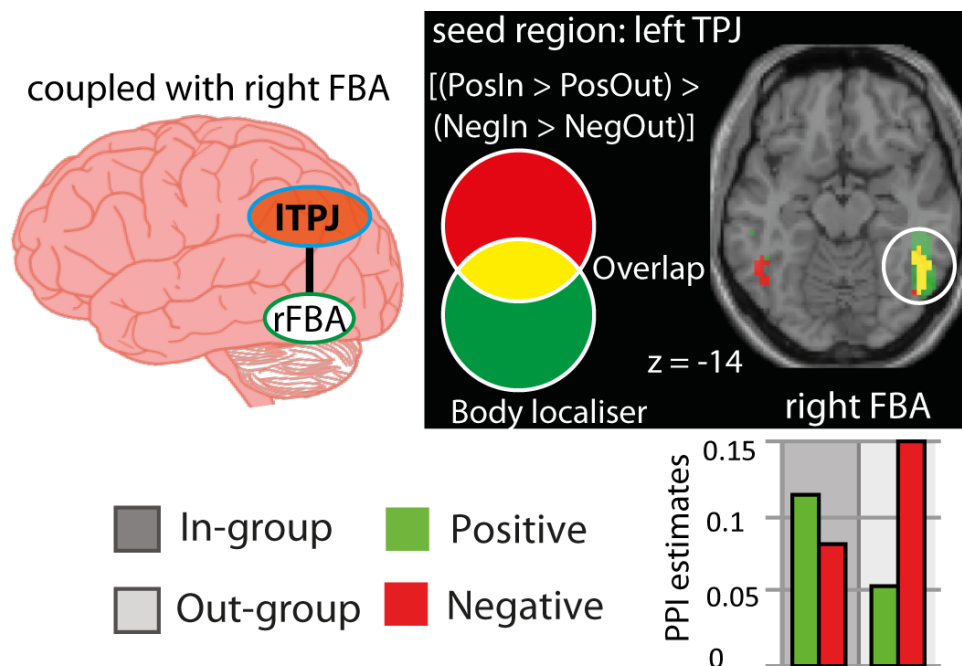
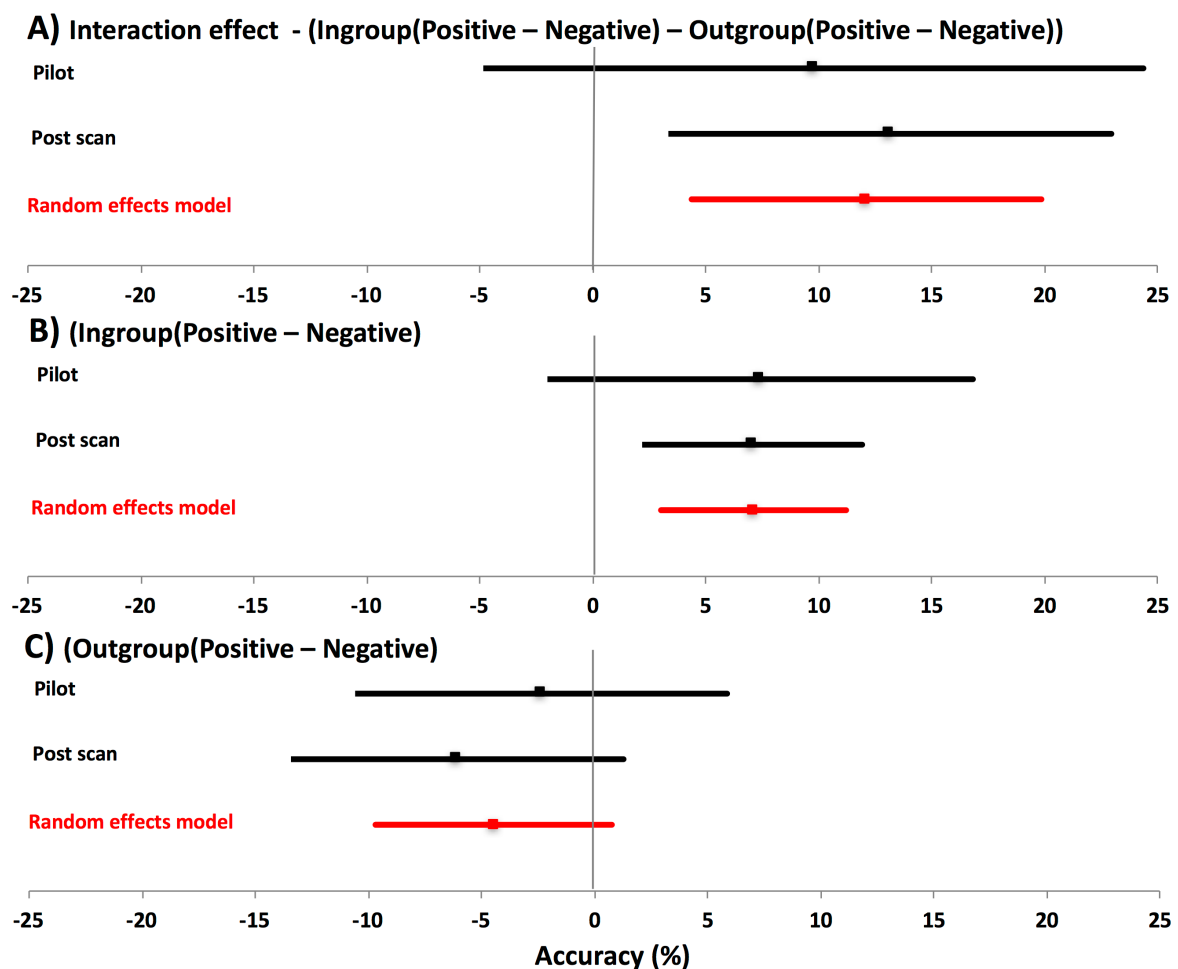


Figure 6. Results from the PsychoPhysiological Interaction (PPI) analysis. Seed regions were identified based on clusters emerging from the Valence by Group [(PosIn>PosOut) > (NegIn > NegOut)] contrast at the univariate level (see Supplementary Tables 1 and 2). Each identified region from the univariate analysis was used as a seed region with the Valence by Group term as the contrast of interest. Clusters emerging from these analyses reveal the strength of correlation over time between activity in that cluster and that in the seed region as a function of the task. PPI analyses revealed that seed region left TPJ (solid orange circle) showed functional coupling with a body-selective patch. A cluster in right FBA showed greater functional coupling with left TPJ when recalling positive and negative traits about in- and out-group members, respectively (shown in red). These areas overlapped with the body-localiser (shown in green; overlap is shown in yellow). The PPI estimates are extracted from a 4 mm sphere around the peak coordinate.

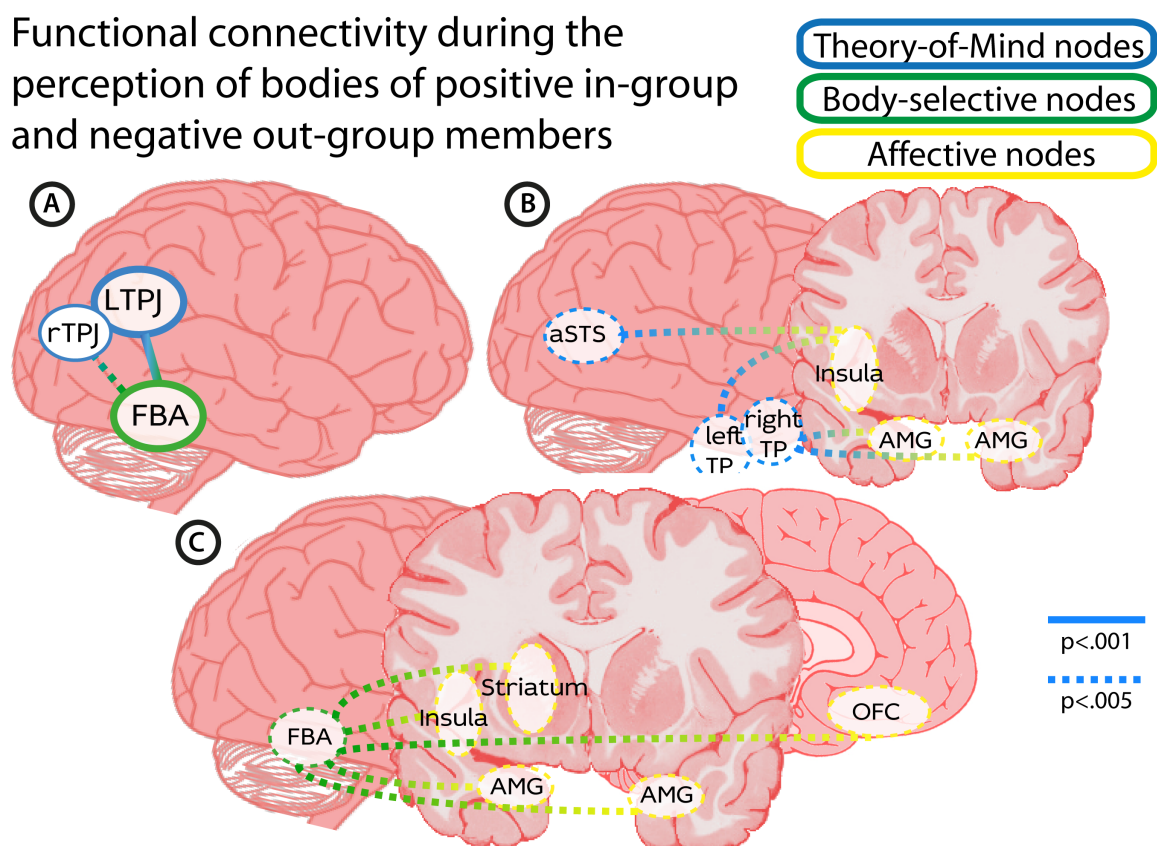


Supplementary Figure 1. Results from a meta-analysis of the pilot and post-scanning behavioural data. Bars represent point estimates and 95% confidence intervals for the effect of interest from each study in the meta-analysis, as well as the combined random effects model.

A) The interaction effect (difference in recall for positive compared to negative trait information for in-group compared to out-group members) shows that the difference in recall accuracy between positive and negative trait information for in-group members is greater than for out-group members. This interaction effect is formed by recall accuracy being greater when recalling positive compared to negative information about B) in-group members and vice versa for C) out-group members. A comparison of B) and C) reveals that the out-group bias for negative trait knowledge is smaller and less consistent than the in-group bias for positive trait knowledge.



Supplementary Figure 2. Summary of the functional connectivity between neural networks involved in body perception (green), Theory of Mind (ToM; blue), and affective processing (yellow) when observing group-members that cued the recall of social knowledge that fit the stereotype (positive in-group and negative out-group members) compared to when it didn't fit the stereotype (negative in-group and positive out-group members). 1) Functional integration of body and ToM networks: right Fusiform Body Area (FBA) is functionally coupled with bilateral TemporoParietal Junction (TPJ). 2) Nodes in the ToM-network couple with the affective network: left temporal pole (TP) couples with left anterior insula, while right temporal pole connects with bilateral amygdala (AMG). Additionally, left posterior insula couples with left anterior Superior Temporal Sulcus (aSTS). 3) Node in the body network couples with the affective network: right FBA couples with bilateral amygdala, left striatum, left amygdala, and the orbitofrontal cortex (OFC). Full line: $p < .001$; Dashed line: $p < .005$.



Supplementary Method

Behavioural pilot experiment

A pilot experiment was completed that had a similar structure to the main experiment. This experiment was conducted to determine whether participants could reliably recognise the encoded bodies after a short retention period.

Method:

Participants

Thirty-one participants (24 females; mean \pm SD age: 20.8 ± 6 years) were recruited from the Bangor community and received course credits for completing the pilot experiment. They gave informed consent according to the local ethics guidelines.

Design

The design was the same as the main experiment, except for the changes below. Participants were evenly divided into two teams (blue and yellow), given a t-shirt of their team's colour to wear, and the two teams completed the experiments in separate rooms. In these experiments, participants observed 128 bodies (64 female of which half were team Blue and half were team Yellow).

Retention period: Following the encoding phase, participants in the pilot experiment would take a short break (~10 min) during which they filled in a questionnaire. This was done to ensure that recency effects (performance on recognition is better for bodies that were presented last during encoding phase; Gershberg and Shimamura, 1994) did not influence the performance on the subsequent recognition test.

Recognition task: On each trial, participants were presented with two bodies (one Blue and one Yellow) and asked to select the body they thought had previously been paired with the simultaneously presented statement. Both of these bodies had previously been presented with positive, negative, or neutral statements. The analyses are the same as for the post-scanning recognition task.

Supplementary Results

Meta-analysis of behavioural data

To provide quantitative support for the pattern of results across both behavioural datasets, we performed a meta-analysis of the pilot and post-scanning data using Exploratory Software for Confidence Intervals (ESCI; Cumming 2012). ESCI weights the contribution of each study as a function of sample size and variability of the estimate to provide a global estimate. Therefore, studies with larger samples and smaller variability have a higher weighting than studies with smaller samples and larger variability.

We meta-analysed three effects of interest (Supplementary Figure 1). The first effect was the interaction term, which was calculated as the difference in recall for positive compared to negative trait information for in-group compared to out-group members. Two further effects comprised the difference in recall performance between positive and negative information for (2) in-group and (3) out-group members separately. We used a random effects model to estimate the size of the effect as recommended by Cumming (2012). Effects are estimated in original units (% accuracy) using 95% confidence intervals. The meta-analysed interaction effect shows that the difference in recall accuracy between positive and negative trait information for in-group members is greater than for out-group members (Supplementary Figure 1A). This interaction effect is formed by recall accuracy being greater when recalling positive compared to negative information about in-group members (Supplementary Figure 1B) and vice versa for out-group members (Supplementary Figure 1C). By comparing Supplementary Figures 1B and 1C, it is clear that the out-group bias for negative trait knowledge is smaller and less consistent than the in-group bias for positive trait knowledge.